

Thermodynamic Differences among Homologous Thermophilic and Mesophilic Proteins[†]

Sandeep Kumar,[‡] Chung-Jung Tsai,[§] and Ruth Nussinov^{*,§,||}

Laboratory of Experimental and Computational Biology and Intramural Research Support Program—SAIC, Laboratory of Experimental and Computational Biology, National Cancer Institute—Frederick, Building 469, Room 151, Frederick, Maryland 21702, and Sackler Institute of Molecular Medicine, Department of Human Genetics and Molecular Medicine, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv 69978, Israel

ABSTRACT: Here, we analyze the thermodynamic parameters and their correlations in families containing homologous thermophilic and mesophilic proteins which show reversible two-state folding \rightleftharpoons unfolding transitions between the native and the denatured states. For the proteins in these families, the melting temperatures correlate with the maximal protein stability change (between the native and the denatured states) as well as with the enthalpic and entropic changes at the melting temperature. In contrast, the heat capacity change is uncorrelated with the melting temperature. These and additional results illustrate that higher melting temperatures are largely obtained via an upshift and broadening of the protein stability curves. Both thermophilic and mesophilic proteins are maximally stable around room temperature. However, the maximal stabilities of thermophilic proteins are considerably greater than those of their mesophilic homologues. At the living temperatures of their respective source organisms, homologous thermophilic and mesophilic proteins have similar stabilities. The protein stability at the living temperature of the source organism does not correlate with the living temperature of the protein. We tie thermodynamic observations to microscopics via the hydrophobic effect and a two-state model of the water structure. We conclude that, to achieve higher stability and greater resistance to high and low temperatures, specific interactions, particularly electrostatic, should be engineered into the protein. The effect of these specific interactions is largely reflected in an increased enthalpy change at the melting temperature.

The thermodynamic stability of a protein may vary with the changes in its environmental conditions (e.g., temperature, pH, buffer, salt concentration, presence and absence of chemical denaturants, concentration of the protein, and presence and absence of substrates, ligands, and subunits). The two most common ways of studying protein stability are via thermal and chemical denaturation, using spectroscopy (circular dichroism (CD) and fluorescence) and differential scanning calorimetry (DSC). Thermal-denaturation experiments often yield three thermodynamic parameters: the melting temperature (T_G), the enthalpy change at the melting temperature (ΔH_G), and the heat capacity change (ΔC_p) between the native (N) and the denatured (D) states of the protein. These parameters can be used to plot the protein

stability curve (1, 2). Protein stability curves describe the temperature-dependent variation of protein stability (the Appendix section gives a description of protein stability curves and the underlying thermodynamic relationships). In studies performed using chemical denaturants such as urea and guanidium hydrochloride (GdnHCl), the Gibbs free energy of unfolding a protein at a given temperature (usually room temperature) is estimated most frequently by using the linear extrapolation method (LEM; 3). LEM gives the so-called m value, the slope of ΔG as a function of the concentration of the denaturant measured around the transition state (3–5).

Protein stability curves illustrate that, for all of the proteins that follow a two-state transition, there are two transition

* To whom correspondence should be addressed at NCI—FCRF Bldg 469, Rm 151, Frederick, MD 21702. Tel: 301-846-5579. Fax: 301-846-5598. E-mail: ruthn@ncifcrf.gov.

[†] The research of R.N. in Israel has been supported in part by Grant No. 95-00208 from BSF, Israel, by a grant from the Israel Science Foundation administered by the Israel Academy of Sciences, by the Magnet grant, by the Ministry of Science grant, and by the Tel Aviv University Basic Research grants and the Center of Excellence administered by the Israel Academy of Sciences. This project has been funded in whole or in part with Federal funds from the National Cancer Institute, National Institutes of Health, under Contract No. NO1-CO-56000.

[‡] Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick.

[§] Intramural Research Support Program—SAIC, Laboratory of Experimental and Computational Biology, National Cancer Institute-Frederick.

^{||} Sackler Institute of Molecular Medicine, Tel Aviv University.

¹ Abbreviations: T_G , heat-denaturation (melting) temperature; T'_G , cold-denaturation temperature; T_L , living temperature of the source organism; T_S , temperature of maximal protein stability; ΔH_G , molar enthalpy change between native and denatured states of a protein at T_G ; ΔS_G , molar entropy change between native and denatured states of a protein at T_G ; ΔC_p , molar heat capacity change between native and denatured states of a protein; $\Delta G(T)$, molar Gibbs free-energy change between the native and denatured states of a protein at temperature T ; $\Delta G(T_S)$, molar Gibbs free-energy change between the native and denatured states of a protein at T_S ; $\Delta G(T_L)$, molar Gibbs free-energy change between the native and denatured states of a protein at T_L ; $\Delta G(298\text{ K})$, molar Gibbs free-energy change between the native and denatured states of a protein at room temperature; EcRnaseH, *Escherichia coli* Ribonuclease H; TtRnaseH, *Thermus thermophilus* Ribonuclease H; °C, degrees Celsius; kcal/mol, kilocalories per mole; Å², square angstroms; DSC, differential scanning calorimetry; CD, circular dichroism; r^2 , square of the linear correlation coefficient.

temperatures where $\Delta G(T) = 0$. These are T_G and T'_G , termed heat- and cold-denaturation transition temperatures, respectively. Protein heat denaturation is driven by its favorable increase in entropy (6), and protein cold denaturation is driven by the favorable decrease in enthalpy (2, 7, 8). The analyses of both ΔC_p and m values have shown that these macroscopic thermodynamic parameters are related to ΔASA , the difference in the accessible surface areas between the denatured and the native protein states (5, 9–11).

To *microscopically* understand the *macroscopic* parameters of protein stability, it is useful to consider a two-state model of the water structure. Such a model enables the prediction of cold denaturation, the presence of molten globule (MG) states in heat but not in cold denaturation, and the hydrophobic effect (Tsai, C.-J., Maizel, J. V., and Nussinov, R., unpublished work). The prediction of cold denaturation is in agreement with the prediction by thermodynamically based stability plots and is consistent with the hydrophobic effect and the minimum solubility of small organic solutes around room temperature (12–15).

Microscopic Scheme. The Gibbs free energy for the cold denatured (D') \rightleftharpoons folded (N) \rightleftharpoons heat denatured (D) protein is defined in terms of enthalpy and entropy changes

$$\Delta G = G_D - G_N = \Delta H - T\Delta S$$

For the denatured states to be favorable at the two extreme temperatures ($T \leq T'_G$ and $T \geq T_G$), the entropy and the enthalpy changes in the system (the protein and the surrounding solvent water) should be compensatory. When proteins denature, they preferentially expose their nonpolar surface area. A nonpolar surface has the effect of ordering the first-shell water molecules, reducing the entropy of the system. A two-state water model is useful in understanding the temperature dependence of protein stability. In such a model, water consists of dynamically transforming different intermolecular hydrogen-bonding types (16, 17). The first is the enthalpically favored, “normal” icelike, tetrahedrally connected hexagonal hydrogen-bonding type, with optimal hydrogen-bond networks. The second is the entropically favorable, enthalpically unfavorable, highly fluctuating liquid form. At low temperatures, the hexagonal ice hydrogen-bonding type prevails. At high temperatures, the denser liquid types dominate, with a fluctuating gradient of interconverting hydrogen-bonding types from low to high temperatures. The water structure is dynamic, with the hydrogen bonds continuously broken and created on a very short time scale (18).

Below the protein cold-denaturation temperature ($T < T'_G$), the fraction of normal hexagonal ice in solvent water increases and the protein cold denatures because of the loss of the hydrophobic effect. In the denatured state of the protein, the exposure of the nonpolar surface promotes optimal-ordered hydrogen-bond formation at the first shell of the water molecules, propagating to the outer shells. This also explains the clathration cagelike formation of small organic solutes in water. The entropy contribution to the Gibbs free-energy change is negative. However, because the hydrogen-bond networks are optimized, the enthalpy is also reduced, overcoming the unfavorable entropy contribution. Hence, at these temperatures, the denatured state (D') of the protein is energetically favorable.

At temperatures above the protein heat-denaturation temperature ($T > T_G$), the protein heat denatures because of an increase in the entropy of the system. Liquid water dominates, with favorable entropy and unfavorable enthalpy terms. Order cannot propagate in highly fluctuating water molecules. The increase in the entropy of the system overcomes the enthalpically unfavorable exposure of the nonpolar surface area of the protein, and the denatured state (D) of the protein is energetically favorable. Hence, at temperatures $T \leq T'_G$ or $T \geq T_G$, the significant differences in the relative fractions of normal hexagonal icelike and highly fluctuating denser liquid hydrogen-bonding types in the surrounding water molecules play critical roles in protein denaturation.

Starting from T_G , as the temperature cools, the fraction of hexagonal ice in solvent water gradually increases, enabling a larger extent of order propagation. This would reduce the entropy of the system. Hence, the denatured state of the protein converts into the more-compact molten-globule (MG) state, burying the nonpolar surface area. The MG state has considerable natively-like secondary structure, a compact fold lacking well-defined tertiary interactions, larger solvent surface accessibility, and low cooperativity of thermal unfolding. Those intermediates not conforming to these characteristics are not included in this definition (19–22). Thus, the D \rightarrow MG step is entropically driven. On the other hand, in the MG \rightarrow N (native-state) folding reaction, the water structure plays an insignificant role. The atomic packing, electrostatic, and disulfide interactions are optimized within the protein. Hence, microscopically, the protein hydrophobic effect is related to the water structure, which is a function of temperature. Macroscopically, it is reflected in ΔASA and is related to ΔC_p . At $T'_G \leq T \leq T_G$, because the protein conformational entropy is reduced in the MG \rightarrow N (6) step and the entropy of the water is reduced in the N \rightarrow D' step, the reduction in the enthalpy of the system should be significant enough to overcome the loss of entropy and drive the reaction to the native state in the first case and to a cold-denatured state in the second. Because MG \rightarrow N is under enthalpy control, the microscopics predict that ΔASA (and ΔC_p) will be uncorrelated with T_G . On the other hand, the formation of specific interactions reflected in ΔH is critically important in heat and cold denaturation.

Macroscopic Analysis: Parameters and Their Correlations. We have collected and analyzed the thermodynamic data relating to families of thermophilic/mesophilic (T/M) proteins. All of these are reversible two-state folding proteins. We present their stability plots and the correlations among their thermodynamic parameters. The melting temperature (T_G), the enthalpy change at the melting temperature (ΔH_G), the entropy change at the melting temperature (ΔS_G), and the maximal protein stability change ($\Delta G(T_S)$) are correlated with one another. These correlations consistently suggest that the higher melting temperatures of thermophilic proteins are obtained by a higher $\Delta G(T_S)$. The increase in $\Delta G(T_S)$ is largely due to the formation of additional specific interactions in the thermophilic proteins as compared to their mesophilic homologues. We find that there is no correlation between ΔC_p and T_G . This observation is particularly interesting because it indicates that an increase in hydrophobicity is not sufficient to improve protein stability. Consistently, we have previously shown that, while different factors contribute to protein thermostability, the higher T_G in thermophiles is best

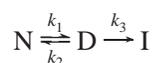
correlated with an increase in factors that promote specific interactions, such as salt bridges and side chain–side chain hydrogen bonds (23). Electrostatic interactions, such as salt bridges and their networks, have favorable electrostatic contributions to the stability of glutamate dehydrogenase from the hyperthermophile *Pyrococcus furiosus* (24, 25).

The proteins in our database show a larger variability in the range of $\Delta G(T_S)$ as compared to $\Delta G(T_L)$, their stabilities at the respective living temperatures of the source organisms. There is no correlation between $\Delta G(T_L)$ and T_L . Specific $\Delta G(T_L)$ values for different proteins may be related to their function. In our dataset, the lowest $\Delta G(T_L)$ is observed for cold-shock proteins and the highest $\Delta G(T_L)$ for archaeal histones, structural proteins responsible for DNA packaging.

MATERIALS AND METHODS

Data Collection and Database Composition. We have performed a literature search using Pubmed to select a database of experimental protein thermodynamic measurements. The search was supplemented by querying the ProTherm database (26). Our aim was to collect thermodynamic data on homologous proteins from (hyper)thermophilic and mesophilic organisms. We have collected five such families. These are archaeal histones (27), SH3 domain-containing proteins (28–31), cold-shock proteins (32–34), Rnase H (35, 36), and the catalytic domains of cellulases (37). These families are termed T/M. The member proteins in the T/M families show high sequence and structural homologies. Together, T/M families contain thermodynamic data on 19 proteins. To compare the results of the analyses on these T/M families, we have also constructed two families of mesophilic proteins. The mesophilic proteins in these two families share sequence or structural homologies to variable extents. Because many of the proteins in these families are only distantly related in sequence, they facilitate the distinguishing between trends related to protein thermostability from phylogenetic differences between thermophiles and mesophiles. We term the mesophilic protein families A/M (all mesophiles). These families consist of five structurally related acylphosphatases (38–40) and the family of Rnase Sa, Sa2, Sa3, Rnase A, Rnase T1, and Barnase (41). The selection of the A/M families for this study is arbitrary. The two A/M families contain 11 proteins. For each protein, we have collected three thermodynamic parameters from the literature: ΔH_G , ΔC_p , and T_G .

For most proteins, thermal denaturation involves some irreversibility. The degree of irreversibility has a greater significance for the thermophilic proteins because the thermolabile amino acids in these proteins may undergo covalent modifications at high temperatures. A complete thermal unfolding process for a two-state protein is described more accurately as follows:



where N, D, and I are the folded native, reversible-denatured, and irreversible-denatured states of the protein, respectively. In thermodynamic experiments, the extent to which protein unfolding follows a two-state mechanism can be measured by the ratio $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$. ΔH^{cal} is the enthalpy change for unfolding determined by DSC. This enthalpy change

value is model-independent. $\Delta H^{\text{van'tHoff}}$ is determined from thermal-denaturation experiments using CD spectroscopy. Alternatively, it is calculated by (42)

$$\Delta H^{\text{van'tHoff}} = 4RT_G^2 C_{p,\text{max}}/\Delta H^{\text{cal}}$$

where $C_{p,\text{max}}$ is the maximum of the excess heat capacity at T_G and R is the universal gas constant. The calculation of the $\Delta H^{\text{van'tHoff}}$ value assumes a two-state folding model. Thus, a value close to unity for $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ indicates the validity of a two-state folding model for a monomeric protein (42, 43). The presence of isodichroic point(s) in the CD spectra recorded at different temperatures in the transition region also indicates that the protein follows a two-state transition. The reversibility of protein unfolding is often measured by the reproducibility of repetitive DSC (or CD) scans on the same protein sample.

We have accepted the claim of the original experimental publications on the reversible two-state nature of the proteins in our database. We have further noted the $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ values and the reversibility of protein folding \rightleftharpoons unfolding transitions wherever available. Further details are given in the original publications.

The accuracy of our analysis depends on the accuracy of the available data. Pace et al. (41) have illustrated that the T_G values are accurate to better than $\pm 1\%$ and that the value for ΔH_G from a van't Hoff analysis can be determined to about $\pm 5\%$, with a good agreement between the different laboratories. However, there may be considerable differences in the ΔC_p values. Recently, Pace and co-workers (41, 44) have noted that the reported values of ΔC_p for Rnase A vary from 1.0 to 2.3 kcal/(mol K). Additionally, it is important to note that, while in general ΔC_p is taken to be independent of T , Privalov (2, 43) has shown that ΔC_p decreases at low and high temperatures.

The number of residues in each protein was used to qualitatively estimate the changes in the accessible surface area (ΔASA) between the native and the denatured states of the protein, using the empirical relationship (9)

$$\Delta\text{ASA} = -907 + 93N_{\text{res}}$$

where N_{res} is the number of residues in a protein. Ideally, we should have measured this quantity from the crystal structures of the proteins. However, high-resolution crystal structures are not available for many of the proteins in our dataset. In those cases where these data are available, the atomic coordinates for several residues are missing. In the following, we present a brief description of each protein family.

Archaeal Histones. The study of Li et al. (27) describes the unfolding of four recombinant archaeal histones: rHFoB from the mesophile *Methanobacterium formicicum*, rHMfA and rHMfB from the hyperthermophile *Methanothermus fervidus*, and rHPyA1 from *Pyrococcus* strain GB-3a. All four of the histones form dimers with two-state ($N_2 \rightleftharpoons 2U$) transitions. The thermodynamic parameters used in this study were reported from thermal-denaturation experiments using DSC and CD spectroscopy. These spectra were recorded at a pH range of 3.0–5.0 in 0.2 M and 1.0 M KCl. The data used here correspond to a pH of 5.0 and to 0.2 M KCl. The thermal unfolding of these proteins is more than 90%

reversible. The values of the $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ ratio for rHmfA and rHFoB are 0.23 and 0.33, respectively. These data are unavailable for other recombinant archaeal histones. Ideally, the value of this ratio should have been 0.5 because the histones are dimers. The archaeal histones family is the only family in our database that contains three thermophiles and one mesophile.

SH3 Domain-Containing Proteins. Data on this family were collected from four different studies (28–31). In total, this family contains data on eight proteins: two (Sac7d and Sso7d) thermophiles and six mesophiles. All eight of the proteins are small single-domain proteins with an SH3 domainlike folding pattern. All of the proteins show two-state folding \rightleftharpoons unfolding transitions. In the case of Sso7d, CD spectra showed the unfolding to be two-state and mostly reversible (28), with the $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ ratio between 0.92 and 0.95. Sac7d also shows reversible unfolding with two-state behavior ($\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}} = 0.97$) at 0.3 M KCl (31). The $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ ratio for the Btk-SH3 domain is ~ 1.1 (30). The thermal unfolding of these proteins has been studied using DSC and CD spectroscopy. The thermodynamic parameters used in this study are from DSC measurements at the neutral pH (7.0). For the mesophilic proteins, the living temperature of the source organisms was taken to be 37 °C.

Cold-Shock Proteins. This family contains data on cold-shock proteins from three organisms: *Escherichia coli* (CspA; 33), *Bacillus subtilis* (CspB; 34), and *Thermotoga maritima* (CspTm; 32). The thermal unfolding of CspTm is $97 \pm 2\%$ reversible, and the $\Delta H^{\text{van'tHoff}}/\Delta H^{\text{cal}}$ ratio is 1.06 ± 0.6 (32). All of the proteins are monomeric and two-state folders. The thermodynamic parameters were collected from DSC studies at the neutral pH (7.0–7.5). This family has two mesophilic and one thermophilic proteins.

Rnase H. The family contains thermodynamic data on the cysteine-free mutants of Rnase H from *E. Coli* (EcRnaseH) and *Thermus thermophilus* (TtRnaseH), reported by Hollien and Marqusee (36) using CD spectroscopy. The authors have also determined the free energies of unfolding from the GdnHCl denaturation of EcRnaseH and TtRnaseH. They found an excellent agreement between the experimentally determined free-energy values and those calculated using the Gibbs–Helmholtz equation for these proteins. This indicates that a two-state folding model is valid for the cysteine-free variants of Rnase H. For the EcRnaseH cysteine-free mutant, the thermal transition is not reversible in the absence of the denaturant. The thermal unfolding curves of both proteins were recorded at a pH of 5.5 in 5 mM NaAc and 50 mM KCl.

Catalytic Domain of Cellulases. This family contains CD measurement data reported by Beadle et al. (37) on the catalytic domains of cellulase E2 from the thermophile *Thermomonospora fusca* (E2_{ca}) and the cellulase CenA from the mesophile *Cellulomonas fimi* (CenA_{P30}). Both of these domains show reversible two-state folding \rightleftharpoons unfolding behavior. For both proteins, the buffer is 50 mM KP_i, 225 mM KCl, and 11.25% ethylene glycol at a pH of 6.8.

Five Structurally Related Proteins. The family contains thermodynamic data on two highly homologous acylphosphatases [muscle (Muscle Acp) and common type (CT-Acp)] reported by Dobson et al. (38, 40). The data on these proteins are compared with three proteins [EcHpr (*E. coli* Hpr), AdB,

and ADA2h] that show structural similarities to muscle Acp (reported in Table 3 of ref 37). All five of the proteins in the family are mesophilic. The living temperature of the source organisms was taken to be 37 °C. All of the proteins show reversible two-state folding \rightleftharpoons unfolding transitions.

Rnase Sa, Sa2, Sa3, Barnase, Rnase T1, and Rnase A. Data on these six ribonucleases were taken from Table 1 in ref 41. All six are mesophilic, and the living temperature of their source organisms was taken to be 37 °C. All of these show reversible two-state folding \rightleftharpoons unfolding transitions. Rnase A is structurally unrelated to Rnase Sa, Sa2, and Sa3 but has a similar function. Rnase Sa, Sa2, and Sa3 show high sequence and structural homologies among themselves. Barnase and Rnase T1 have low sequence and structural similarities with these proteins. The values of the $\Delta H^{\text{cal}}/\Delta H^{\text{van'tHoff}}$ ratios are 0.99, 1.00, and 0.96 for Rnase Sa, Sa2, and Sa3, respectively. The thermal denaturation for these proteins is more than 95% reversible (41).

The data in the literature are frequently reported in SI and non-SI units. For uniformity, we used the unit for energy as calories. The conversion factor between calories and joules is 1 cal = 4.184 J.

Computation of Linear Correlation Coefficients and *t* Values. For each pair of thermodynamic parameters (x, y : $x_1, y_1, x_2, y_2, \dots, x_n, y_n$) in our database, we have fitted a line, $y = a + bx$, using the least-squares procedure. The linear correlation coefficient is calculated by

$$r = \frac{(n \sum xy - \sum x \sum y)}{\sqrt{(n \sum x^2 - (\sum x)^2)(n \sum y^2 - (\sum y)^2)}}$$

Our dataset can be regarded as a sample of protein populations. In this sense, the sampling theory of correlation can be used to determine if the correlations observed in our dataset are relevant for proteins in general. We formulate the null hypothesis that the population correlation coefficient (ρ) for a given parameter pair is zero (H_0 ; $\rho = 0$), while the linear correlation coefficient for the same parameter pair is r in our dataset. The t value is computed to test the null hypothesis by

$$t = r\sqrt{n-2}/\sqrt{1-r^2}$$

where n is the number of proteins in our dataset. For proteins in the T/M families, the null hypothesis is rejected at the 99% level of confidence if $t > 2.60$ (45). The rejection of the null hypothesis for a parameter pair indicates that the two parameters are likely to be correlated with each other in proteins.

RESULTS

Thermodynamic Parameters and Protein Stability Curves for Thermophiles and Mesophiles. Parts a–e of Figure 1 present the stability plots for five homologous T/M protein families. Parts a and b of Figure 2 present the stability plots for two mesophilic (A/M) protein families. Table 1a details the corresponding protein families, their sizes, their ΔASA , and the thermodynamic values collected from the literature. Table 1b lists the temperature of maximal protein stability (T_s), the Gibbs free-energy change for protein unfolding at T_s ($\Delta G(T_s)$), the living temperature of the source organism

Table 1: Structural and Thermodynamic Parameters for Proteins in Our Database and Parameters Derived from Protein Stability Curves

a. Structural and Thermodynamic Parameters for Proteins ^a											
protein name	PDB file	N_{res}	ΔASA (\AA^2)	ΔH_G (kcal/mol)	ΔC_p (kcal/(mol K))	T_G ($^{\circ}\text{C}$)					
Archaeal Histones ^b											
rHfob		134	11555	115.9 \pm 0.1	2.6	74.8 \pm 0.2					
rHMfA	1B67	136	11741	164.1 \pm 1.3	2.2	104.1 \pm 0.3					
rHMfB		138	11927	150.3 \pm 0.9	1.9	112.8 \pm 0.3					
rHPyA1		134	11555	184.4 \pm 3.0	2.4 \pm 0.2	114.1 \pm 0.6					
SH3 Domain-Containing Proteins ^c											
Itk		60	4673	42.5	0.8 \pm 0.1	69					
Tec		66	5231	40.4	0.7	71					
Btk		67	5324	46.8	0.7 \pm 0.1	80					
α -Spectrin	1SHG	62	4859	47.1 \pm 2.4	0.8	66 \pm 0.2					
Abl	1ABQ	63	4952	46.4 \pm 2.4	0.8 \pm 0.1	68.5 \pm 0.2					
Fyn		64	5045	55.7 \pm 3.6	0.8 \pm 0.1	70.6 \pm 0.2					
Sac7d	1AZP	66	5231	60.2	0.9	90.7					
Sso7d	1BF4	64	5045	63.3	0.6	98.5					
Cold-Shock Proteins ^d											
CspA	1MJC	70	5603	43.3 \pm 1.7	0.8	57					
CspB	1CSP	67	5324	36.8	0.9 \pm 0.2	53.4					
CspTm		66	5231	62.6 \pm 3.1	1.1 \pm 0.1	82 \pm 0.2					
Cysteine-Free Mutants of RNase H ^e											
EcRNaseH	2RN2	155	13508	120 \pm 4	2.7 \pm 0.2	66 \pm 1					
TiRNaseH	1RIL	166	14531	131 \pm 5	1.8 \pm 0.1	86 \pm 1					
Catalytic Domains of Cellulases ^f											
CenA _{p30}				107 \pm 3.1	3.8	56.4 \pm 0.3					
E2 _{cd}				190 \pm 14	3.8	72.2 \pm 0.2					
Five Structurally Related Proteins ^g											
muscle Acp		98	8207	93.5 \pm 4.7	1.5 \pm 0.2	56.7 \pm 2.0					
CT-Acp	2ACY	98	8207	69.3 \pm 3.5	1.5 \pm 0.2	54.0 \pm 2.0					
EcHpr	1POH	85	6998	75.8 \pm 1.2	1.5 \pm 0.1	63.6 \pm 0.1					
AdB		82	6719	71.5	0.9	74.2					
ADA2h		80	6533	47.6	0.9	77					
RNase Sa, Sa2, Sa3, Barnase, RNase T1, and RNase A ^h											
RNase Sa	1RGG	96	8021	97.4 \pm 4.9	1.5 \pm 0.1	48.4 \pm 0.3					
RNase Sa2		97	8114	68.4 \pm 3.4	1.3 \pm 0.1	41.1 \pm 0.3					
RNase Sa3		99	8300	93.6 \pm 4.7	1.6 \pm 0.1	47.2 \pm 0.3					
RNase T1	9RNT	104	8765	105.7	1.7	51.6					
Barnase	1RNB	110	9323	126.6	1.8	53.2					
RNase A	1AFK	124	10625	119.4	1.9	62.8					
b. Parameters Derived from Protein Stability Curves ⁱ											
protein name	T_S ($^{\circ}\text{C}$)	$\Delta G(T_S)$ (kcal/mol)	T_L ($^{\circ}\text{C}$)	$\Delta G(T_L)$ (kcal/mol)	$\Delta G(298\text{ K})$ (kcal/mol)	protein name	T_S ($^{\circ}\text{C}$)	$\Delta G(T_S)$ (kcal/mol)	T_L ($^{\circ}\text{C}$)	$\Delta G(T_L)$ (kcal/mol)	$\Delta G(298\text{ K})$ (kcal/mol)
Archaeal Histones											
rHfob	32.2	7.2	43.0	6.8	7.0	rHMfB	40.2	14.6	83.0	9.4	13.9
rHMfA	35.3	15.5	83.0	7.9	15.1	rHPyA1	44.1	17.2	95.0	8.0	15.8
SH3 Domain-Containing Proteins											
Itk	20.5	3.1	37.0	2.7	3.1	Abl	14.5	3.8	37.0	3.1	3.6
Tec	17.4	3.2	37.0	2.8	3.2	Fyn	6.8	5.3	37.0	4.1	4.9
Btk	22.2	3.9	37.0	3.7	3.9	Sac7d	26.9	5.4	80.0	1.6	5.4
α -Spectrin	12.7	3.8	37.0	3.0	3.6	Sso7d	9.2	8.0	77.0	3.3	7.7
Cold-Shock Proteins											
CspA	5.0	3.5	37.0	2.2	3.0	CspTm	29.4	4.8	80.0	0.3	4.7
CspB	14.3	2.3	37.0	1.5	2.1						
Cysteine-Free Mutants of RNase H											
EcRNaseH	24.3	7.5	37.0	6.8	7.5	TiRNaseH	20.1	12.4	68.5	5.6	12.4
Catalytic Domains of Cellulases											
CenA _{p30}	29.4	4.4	37.0	4.1	4.3	E2 _{cd}	25.7	13.1	55.0	7.8	13.1
Five Structurally Related Proteins											
muscle Acp	0.9	8.2	37.0	4.7	6.6	AdB	5.2	7.4	37.0	5.8	6.7
CT-Acp	9.8	4.8	37.0	2.9	4.2	ADA2h	25.8	3.6	37.0	3.4	3.6
EcHpr	16.4	5.4	37.0	4.4	5.3						
RNase Sa, Sa2, Sa3, Barnase, RNase T1, and RNase A											
RNase Sa	-9.7	9.1	37.0	3.1	5.8	RNase T1	-6.5	9.8	37.0	4.2	6.8
RNase Sa2	-8.4	5.5	37.0	0.9	3.0	Barnase	-10.4	12.8	37.0	5.6	8.7
RNase Sa3	-7.2	8.2	37.0	2.7	5.3	RNase A	5.5	10.5	37.0	7.2	9.3

^a The thermodynamic parameters ΔH_G , ΔC_p , and T_G were obtained from the literature. N_{res} denotes the number of residues in a protein. ΔASA is the change in accessible surface area between the native and the denatured states of the protein. ΔH_G and ΔC_p are the change in enthalpy at the melting temperature (T_G) and the change in heat capacity between the native and the denatured states, respectively. The availability of X-ray crystal structure data for the proteins in our database is indicated by their protein data bank (PDB) file names. ^b Data from ref 27. ^c Data from refs 28–31.

^d Data from refs 32–34. ^e Data from refs 35 and 36. ^f Data from ref 37. ^g Data from refs 38–40. ^h Data from ref 41. ⁱ The thermodynamic parameters, except T_L , were calculated using eqs 1, 6, and 7 in the Appendix section. T_S is the temperature at which the protein is maximally stable, and $\Delta G(T_S)$ is the maximal Gibbs free-energy change of unfolding for the proteins. T_L is the living temperature of the source organism, and $\Delta G(T_L)$ is the Gibbs free-energy change of unfolding at the living temperature of the source organism. $\Delta G(298\text{ K})$ is the Gibbs free-energy change for protein unfolding at 298 K (room temperature).

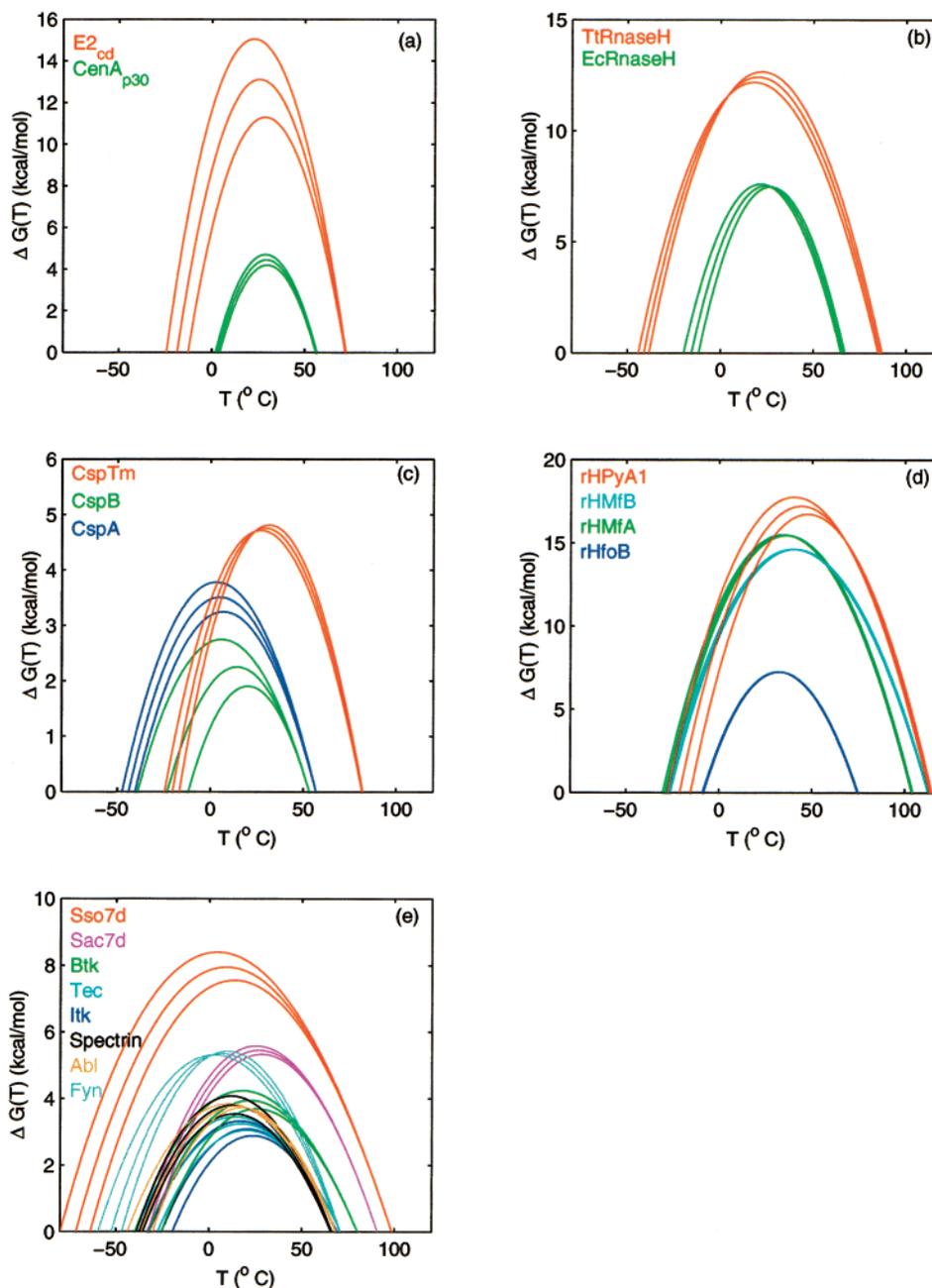


FIGURE 1: Protein stability curves for proteins in T/M (thermophile/mesophile) families. Catalytic domains in (a) cellulases, (b) Rnase H, (c) cold-shock proteins, (d) archaeal histones, and (e) SH3 domain-containing proteins. All of the proteins within a T/M family show a high degree of sequence and structural similarities. In each plot, the X axis represents the temperature and the Y axis represents the change in Gibbs free energy for protein unfolding ($\Delta G(T)$). The errors in $\Delta G(T)$ were estimated using the available experimental errors in ΔH_G , ΔC_p , and T_G for each protein.

(T_L), the Gibbs free-energy change for protein unfolding at T_L ($\Delta G(T_L)$), and the Gibbs free-energy change for protein unfolding at room temperature ($\Delta G(298\text{ K})$) calculated from the protein stability curves plotted using the Gibbs–Helmholtz equation (see the Appendix section).

In the T/M families, greater protein stability and resistance to higher temperatures are obtained by an upshift and broadening of their protein stability curves as compared to their mesophilic homologues (Figure 1). In most cases, the estimated cold-denaturation temperatures (T'_G) are also lower for the thermophilic proteins. The exception to these observations is the cold-shock protein family, where the protein stability curve for the thermophilic (*T. maritima*) cold-shock protein (CspTm) is both upshifted and right-

shifted. These proteins are only marginally stable at the living temperatures of the source organism (Table 1b). For example, $\Delta G(T_L)$ for CspTm is only 0.35 kcal/mol.

Mesophilic proteins do not show consistent trends (Figure 2). In the A/M family containing ribonucleases, the difference in the maximal stabilities of Barnase and Rnase Sa2 is ~ 7 kcal/mol (Figure 2b and Table 1b). The difference in their melting temperatures is 12.2 °C, and Barnase contains 13 more amino acids than Rnase Sa2. Barnase is only distantly related to Rnase Sa2 (41). Hence, the differences in stability for these proteins may be due to phylogeny rather than to thermophilicity.

A higher melting (transition) temperature in a thermophilic protein can be attained in one of three ways (31, 37, 46,

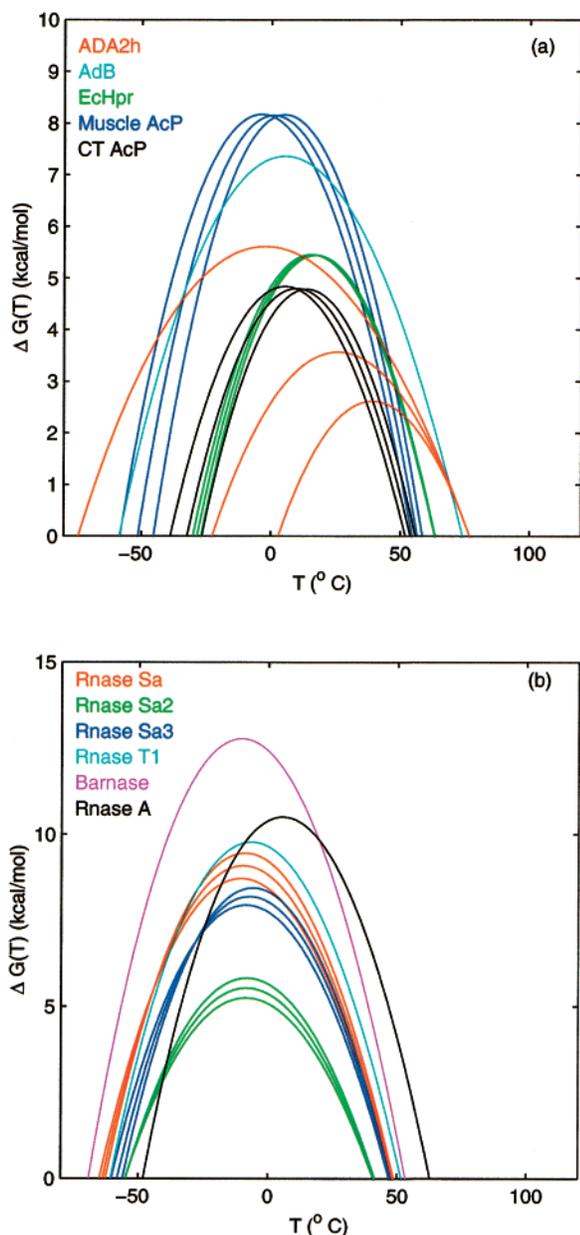


FIGURE 2: Protein stability curves for proteins in A/M (all mesophile) families: (a) muscle and common type Acp and their structural homologues and (b) Rnase Sa, Sa2, Sa3, Barnase, Rnase A, and Rnase T1. Rnase Sa, Sa2, and Sa3 show high sequence similarities. Barnase, Rnase T1, and Rnase A are related only distantly to Rnase Sa.

47). First, the maximal protein stability, $\Delta G(T_S)$, could be larger for the thermophile as compared to its homologous mesophile, resulting in an upshift of the curve and hence a higher T_G . Such an occurrence would be the outcome of a larger ΔH_G , which is responsible for the slope of the curve at T_G . Alternatively, the $\Delta G(T_S)$ values of both proteins might be similar; however, their curvatures may differ. If the curvature (specified largely by ΔC_p) is smaller, the curve would be broader, leading to the same eventual observation. Third, we could observe a left/right shift of one curve with respect to the other. The crucial question is whether the hyperthermophilic proteins have a greater maximal protein stability ($\Delta G(T_S)$) as compared to their mesophilic homologues (47). Figure 1 illustrates that, in most cases, if T_G is higher, the maximal protein stability $\Delta G(T_S)$ is larger too,

Table 2: Regression Analysis among Various Thermodynamic Parameters for Proteins in T/M Families^a

parameter pair	r^2	t value	parameter pair	r^2	t value
$\Delta H_G, T_G$	0.54	4.20	$\Delta C_p, \Delta G(T_S)$	0.51	3.95
$\Delta h^\ddagger_G, T_G$	0.81	7.90	$\Delta C_p, \Delta ASA$	0.85	9.22
$\Delta H_G, T_S$	0.62	4.95	$\Delta G(T_S), T_G$	0.73	6.37
$\Delta H_G, \Delta C_p$	0.77	7.09	$\Delta g^\ddagger(T_S), T_G$	0.83	8.55
$\Delta H_G, \Delta G(T_S)$	0.92	13.13	$\Delta G(T_S), T_S$	0.52	4.03
$\Delta h^\ddagger_G, \Delta g^\ddagger(T_S)$	0.82	8.34	$\Delta G(T_S), \Delta ASA$	0.62	4.95
$\Delta H_G, \Delta ASA$	0.80	7.75	$\Delta G(T_S), \Delta G(T_L)$	0.71	6.06
$\Delta H_G, N_{res}$	0.80	7.75	$\Delta G(T_S), T_L$	0.59	4.65
$\Delta s^\ddagger_G, T_G$	0.72	6.15	$\Delta G(T_L), T_L$	0.13	1.50
$\Delta S_G, \Delta G(T_S)$	0.88	10.34	$T_G, \Delta ASA$	0.20	1.94
$\Delta s^\ddagger_G, \Delta g^\ddagger(T_S)$	0.77	7.13	T_G, N_{res}	0.20	1.97
$\Delta C_p, T_G$	0.16	1.69	$T_S, \Delta ASA$	0.38	3.03
$\Delta C_p, T_S$	0.52	4.03	T_G, T_S	0.52	4.03

^a r^2 denotes the square of the linear correlation coefficient for the parameter pairs in our dataset. The t value was computed to test the null hypothesis that the population correlation coefficient (ρ) for the parameter pair is zero ($H_0: \rho = 0$; Materials and Methods section). For the proteins in the T/M families, the null hypothesis is rejected at the 99% level of confidence if $t > 2.60$. Note that the null hypothesis is accepted for the parameter pairs ($\Delta C_p, T_G$), ($T_G, \Delta ASA$), (T_G, N_{res}), and ($\Delta G(T_L), T_L$) for the proteins in T/M families. Hence, these parameters are unlikely to be correlated in proteins. The parameter pairs for which t values are insignificant are indicated in bold. Superscript \ddagger shows the normalized values of enthalpy and entropy change at the melting temperature T_G (Δh^\ddagger_G and Δs^\ddagger_G), and superscript \ddagger shows the normalized maximal protein stability change ($\Delta g^\ddagger(T_S)$). For each protein, the values of Δh^\ddagger_G , Δs^\ddagger_G , and $\Delta g^\ddagger(T_S)$ were computed by dividing ΔH_G , ΔS_G , and $\Delta G(T_S)$, respectively, by the number of residues in the protein (N_{res}). For each protein, $\Delta S_G = \Delta H_G/T_G$.

indicating that, at least in this limited T/M dataset, a higher T_G is reached via an upshift of the stability curves. Recently, Rees and Robertson (48) have also reported similar observations. Table 1b shows that thermophilic proteins have a greater $\Delta G(T_S)$ and $\Delta G(298\text{ K})$ values than their mesophilic homologues.

Within the T/M families, regardless of the heat transition temperatures (T_G), the temperature of maximal stability (T_S) falls frequently around room temperature. The average temperature of maximal stability for the 19 thermophilic and mesophilic proteins is 22.6 ± 11.0 °C. For the eight thermophilic proteins, rHMfA, rHMfB, rHPyA1, Sac7d, Sso7d, CspTm, TtRnaseH, and E2_{cd}, the average temperature of maximal stability is 28.9 ± 11.2 °C. For the 11 mesophilic homologues of these 8 thermophiles, the average temperature of maximal protein stability is 18.1 ± 8.6 °C. This temperature range coincides with room temperature (25 °C). At room temperature, the solubility of small organic solutes is minimal (e.g., the solubility of hexane in water at 25 °C is only 2×10^{-6} mole fractions (15)), pointing to the hydrophobic effect (12–15).

Correlations among Various Thermodynamic Parameters. The squares of the linear correlation coefficients obtained by the regression analyses of various thermodynamic parameters of T/M families are given in Table 2. One T/M family, the catalytic domain of cellulases, has been excluded from the analysis. In this case, the original experimental report (37) does not describe the determination of ΔC_p for these domains. Instead, the authors have set the value of ΔC_p at 3.8 kcal/(mol K) for both thermophilic and mesophilic enzymes on the basis of theoretical considerations of their size and hydrophobicity. Hence, our correlation analyses

Table 3: Comparison between Average Values for Protein Stabilities at the Living Temperatures of Source Organisms, at Room Temperature, and at Maximal Protein Stabilities^a

protein family	average T_L (°C)	$\Delta G(T_L)$ (kcal/mol)	$\Delta G(298\text{ K})$ (kcal/mol)	$\Delta G(298\text{ K})$ (kcal/mol)
archaeal histones	76	8.0 ± 1.1	13.6 ± 4.4	13.0 ± 4.0
SH3 domain-containing proteins	47	3.0 ± 0.7	4.6 ± 1.6	4.4 ± 1.6
cold-shock proteins	51	1.3 ± 0.9	3.5 ± 1.3	3.3 ± 1.3
Rnase H	53	6.2	10.0	9.9
cellulases	46	6.0	8.8	8.7
thermophilic proteins	78	5.5 ± 3.4	11.4 ± 4.7	11.0 ± 4.4
mesophilic proteins ^b	37	3.7 ± 1.7	4.4 ± 1.7	4.2 ± 1.7

^a The average T_L is calculated by taking the average of the living temperatures of the organisms containing the proteins. There is no correlation between $\Delta G(T_L)$ and T_L . ^b The mesophilic proteins in the T/M families.

involve 17 of the 19 proteins in the T/M families. Qualitatively, the correlations may be divided into different categories on the basis of the square of the linear correlation coefficient (r^2) values. We consider a correlation between two parameters to be nonexistent if $r^2 < 0.3$, weak if $0.3 \leq r^2 < 0.6$, good if $0.6 \leq r^2 < 0.8$, and very good if $r^2 \geq 0.8$. The significance of the observed correlations is measured by t statistics.

The observations noted in Table 2 yield a consistent picture. No correlation is observed between ΔC_p and T_G ($r^2 = 0.16$). ΔC_p is correlated with ΔASA , the change in the accessible surface area of the protein upon folding, and the number of residues (N_{res}) in the protein (5, 9–11). ΔASA ($ASA_{\text{denatured}} - ASA_{\text{native}}$) is largely nonpolar. During protein folding, most (~80%) of the nonpolar surface area is buried (23). Hence, the lack of correlation between ΔC_p and the heat transition temperature T_G indicates that the melting temperature is unrelated to the extent of exposure of the nonpolar surface area upon denaturation. Further, T_G does not correlate with either ΔASA or N_{res} . The t values for correlations between T_G and ΔC_p , between T_G and ΔASA , and between T_G and N_{res} are insignificant (shown in bold in Table 2). The normalization of ΔC_p by the number of residues in the proteins (N_{res}) does not yield a better correlation with T_G ($r^2 = 0.04$). Hence, these parameters are likely to be uncorrelated in proteins, implying that a higher T_G is not obtained solely by strengthening the hydrophobic effect. Very good correlations are observed between $\Delta G(T_S)$ and ΔH_G ($r^2 = 0.92$) and between $\Delta G(T_S)$ and ΔS_G ($\Delta S_G = \Delta H_G/T_G$; $r^2 = 0.88$). The r^2 value between $\Delta g(T_S)$ and Δh_G is 0.82, and between $\Delta g(T_S)$ and Δs_G it is 0.77 (Table 2). For each protein, Δh_G , Δs_G , and $\Delta g(T_S)$ values are obtained by normalizing ΔH_G , ΔS_G , and $\Delta G(T_S)$, respectively, by N_{res} . These observations indicate that a greater maximal protein stability is obtained by the greater enthalpic and entropic changes at the melting temperature. Consistently, a good correlation is observed between $\Delta G(T_S)$ and T_G ($r^2 = 0.73$). Normalizing $\Delta G(T_S)$ by N_{res} yields a very good correlation between $\Delta g(T_S)$ and T_G ($r^2 = 0.83$). T_G also correlates well with both Δh_G ($r^2 = 0.81$) and Δs_G ($r^2 = 0.72$). Figure 3 presents the linear regression plots for five thermodynamic parameter pairs for T/M families, namely, Δh_G versus T_G , Δs_G versus T_G , $\Delta g(T_S)$ versus Δh_G , $\Delta g(T_S)$ versus Δs_G , and $\Delta g(T_S)$ versus T_G . The t values for these parameter pairs are highly significant at the 99% level of confidence, indicating that these parameters are likely to be correlated in proteins. These observations illustrate that the higher melting temperatures of thermophilic proteins are the outcome of specific additional interactions, such as electrostatics, in addition to the nonpolar buried surface area. The contribution of the

hydrophobic effect ($\Delta G(T_S)$ versus ΔASA is $r^2 \sim 0.6$; Table 2) is to a lesser extent. Several of the other correlations listed in Table 2 are simply the outcome of the functional relationships among various thermodynamic parameters. For example, good correlations of ΔH_G with ΔC_p , ΔASA , and N_{res} are not surprising. The observed good correlations among T_G , ΔH_G , ΔS_G , and $\Delta G(T_S)$ are also expected.

Living Temperature, Protein Stability, and Function. Tables 2 and 3 list a number of interesting observations regarding protein stability and function. First, $\Delta G(T_L)$, the stability of the protein at the living temperature of the source organism is uncorrelated with the living temperature, T_L ($r^2 \sim 0.1$). t statistics also show that these parameters are likely to be uncorrelated in proteins (Table 2). Second, within each protein family, $\Delta G(T_L)$ is relatively constant. For example, the stabilities of archaeal hyperthermophilic histone rHPyA1 and its mesophilic homologue rHFoB at the respective living temperatures of their source organisms differ by only 1.2 kcal/mol. In contrast, the maximal protein stabilities of these two histones differ by 10 kcal/mol (Table 1b). These observations indicate that the stability of individual proteins in the source organism may relate to their function or to molecular events which relate to their function. Proteins are known to be flexible and often undergo conformational changes to perform their function (49). Too high of a stability may hinder protein function. Consistently, the maximal stabilities of proteins in our database are at temperatures lower than the organism living temperatures (Table 1b). On average, $\Delta G(T_L)$ is approximately 3 kcal/mol lower than $\Delta G(T_S)$ (Table 3). Consideration of the protein function may also explain the left/right shift of the stability curves of cold-shock proteins. Similarly, archaeal histones, structural proteins responsible for DNA packaging, have the highest $\Delta G(T_S)$ and $\Delta G(T_L)$ values in our database.

Interestingly, there is a good correlation between $\Delta G(T_S)$ and $\Delta G(T_L)$ ($r^2 = 0.71$) in T/M families. Thus, higher maximal stabilities may imply marginally higher stabilities at the living temperatures. Five of the eight thermophilic proteins in our database, rHMfA, rHMfB, rHPyA1, Sso7d, and E2_{cd}, have higher stabilities at their living temperatures as compared to their mesophilic homologues (Table 1b). The remaining three thermophilic proteins, Sac7d, CspTm, and TtRnaseH, are less stable at their source-organism living temperatures than their mesophilic homologues. It is interesting to compare the mean values of $\Delta G(T_S)$ and $\Delta G(T_L)$ of the thermophiles with those of the mesophiles. In the thermophiles, $\Delta G(T_S)_{\text{ave}} = 11.4 \pm 4.7$ kcal/mol and $\Delta G(T_L)_{\text{ave}} = 5.5 \pm 3.4$ kcal/mol. In the mesophiles, $\Delta G(T_S)_{\text{ave}} = 4.4 \pm 1.7$ kcal/mol and $\Delta G(T_L)_{\text{ave}} = 3.7 \pm 1.7$ kcal/mol. Thus, while there is a 7 kcal/mol difference in the maximal

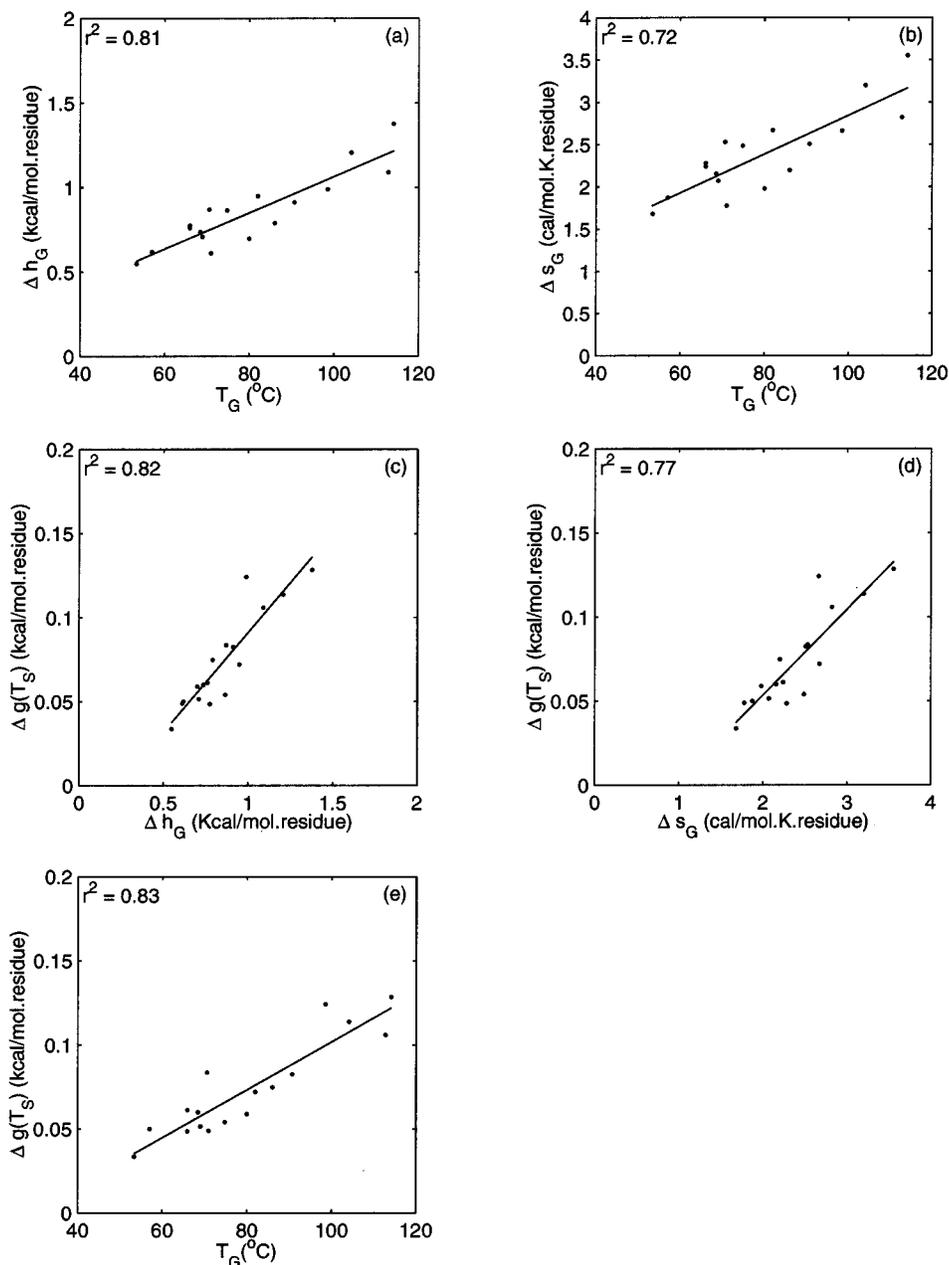


FIGURE 3: Regression analysis using line fitting by least squares for (a) Δh_G versus T_G , (b) Δs_G versus T_G , (c) $\Delta g(T_S)$ versus Δh_G , (d) $\Delta g(T_S)$ versus Δs_G , and (e) $\Delta g(T_S)$ versus T_G . In each plot, the square of the linear correlation coefficient (r^2) is presented in the top left corner. For each protein, Δh_G , Δs_G , and $\Delta g(T_S)$ values were obtained by dividing the ΔH_G , ΔS_G , and $\Delta G(T_S)$ values of the protein by the number of residues in the protein (N_{res}), respectively. For each protein, $\Delta s_G = \Delta H_G/T_G$.

protein stabilities, there is only a 1.8 kcal/mol difference in the stabilities at the corresponding living temperatures between the thermophiles and the mesophiles. Similarly, at room temperature, the thermophilic proteins in our database have $\Delta G(298\text{ K})_{ave} = 11.0 \pm 4.4$ kcal/mol, while their mesophilic homologues have $\Delta G(298\text{ K})_{ave} = 4.2 \pm 1.7$ kcal/mol (Table 3). These results appear reasonable. Thermophilic proteins have evolved to perform optimally at the relevant living temperatures of the thermophilic organisms.

ΔH_G , ΔC_p , and Protein Stability. ΔC_p determines the curvature and ΔH_G the slope of the protein stability curve at T_G . A lower ΔC_p value will broaden and an increased ΔH_G will raise the protein stability curve. Both the lowering of ΔC_p and the increasing of ΔH_G result in a higher $\Delta G(T_S)$ (eq 7 in the Appendix section). For the thermophilic proteins in our database, a higher T_G is achieved by a higher $\Delta G(T_S)$.

Five out of the eight thermophilic proteins, rHMfA, rHMfB, rHPyA1, Sso7d, and TtRnaseH, have higher ΔH_G and lower ΔC_p values than their mesophilic homologues (Table 1a). Three proteins do not follow this trend: Sac7d, CspTm, and E2cd. The ΔC_p value for Sac7d is comparable to those of mesophilic SH3 domains. However, the difference between the T_G values of Sac7d and other SH3 domain-containing proteins is not so large, and $\Delta G(T_S)$ for Sac7d is comparable to those of other SH3 domain proteins (Table 1b and Figure 1e). For E2cd, ΔC_p has not been determined experimentally (37). For CspTm, the value of $\Delta C_p = 1.1 \pm 0.1$ kcal/(mol K) was obtained from the slope of ΔH^{cal} versus the transition temperature (T_{tr}) in the 5.0–8.5 pH range (32). Here, we have used this value of ΔC_p for CspTm because the values determined this way are considered to be more reliable. However, we note that $\Delta C_p = 0.36$ kcal/(mol K) in the

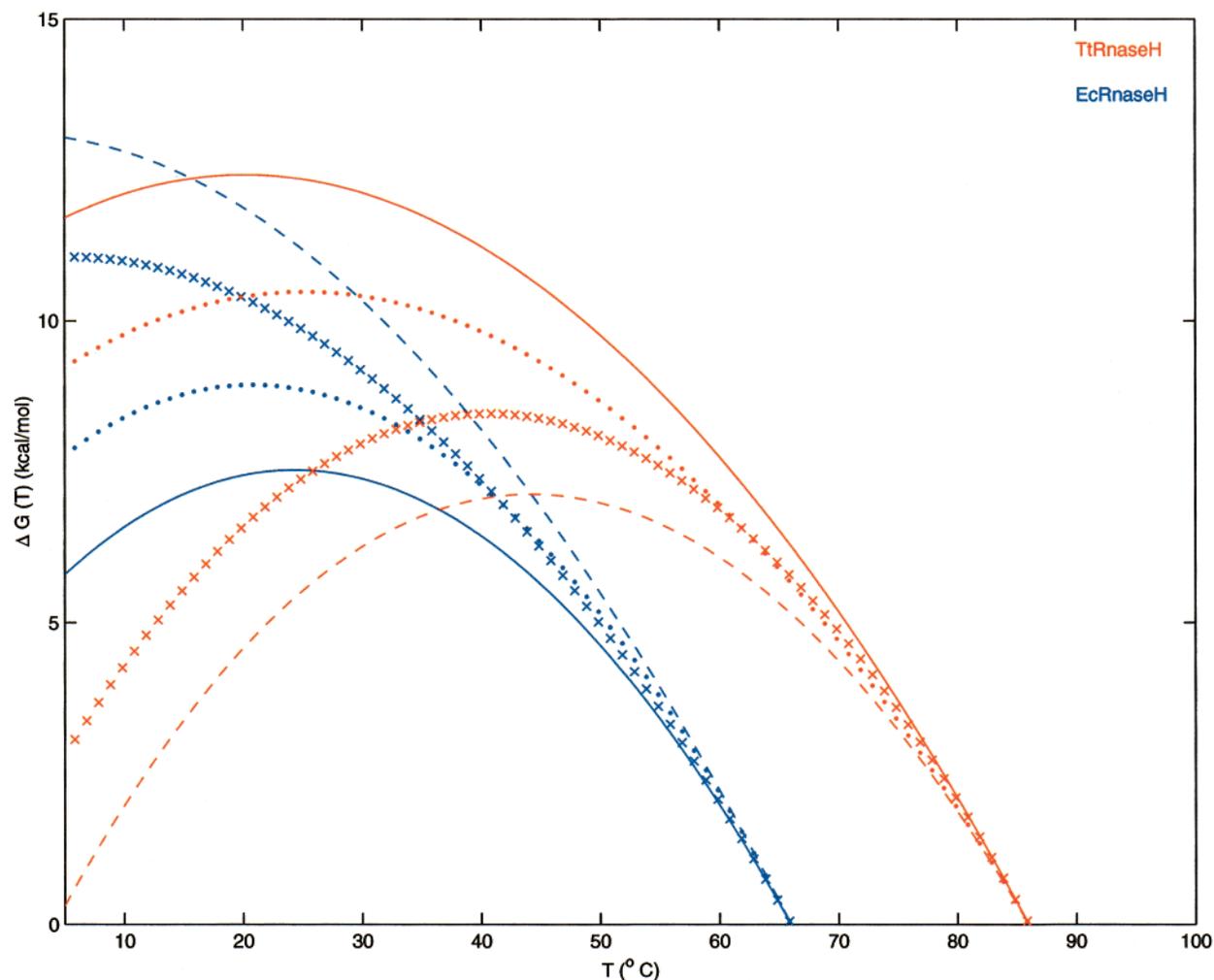


FIGURE 4: Effect of interchanging thermodynamic parameters on the stability curves of thermophilic (TtRnaseH, red color) and mesophilic (EcRnaseH, blue color) ribonuclease H. The X axis represents temperature and the Y axis the change in $\Delta G(T)$ in kcal/mol. In each case, the solid curve represents the original data, the dotted curve represents the effect of interchanging ΔH_G , the \times -marked curve represents the effect of interchanging ΔC_p , and the dashed curve represents the effect of interchanging both ΔH_G and ΔC_p .

Table 4: Effect of Interchanging Thermodynamic Parameters on Thermophilic and Mesophilic Ribonuclease H^a

protein	swapped parameter	ΔH_G (kcal/mol)	ΔC_p (kcal/(mol K))	T_G (°C)	T_S (°C)	$\Delta G(T_S)$ (kcal/mol)	T_L (°C)	$\Delta G(T_L)$ (kcal/mol)	$\Delta G(298\text{ K})$ (kcal/mol)
EcRnaseH	none	120.0	2.7	66.0	24.3	7.5	37.0	6.8	7.5
EcRnaseH	ΔH_G	131.0	2.7	66.0	20.8	8.9	37.0	7.8	8.9
EcRnaseH	ΔC_p	120.0	1.8	66.0	5.5	11.1	37.0	8.0	9.9
EcRnaseH	$\Delta H_G, \Delta C_p$	131.0	1.8	66.0	0.5	13.1	37.0	8.9	11.2
TtRnaseH	none	131.0	1.8	86.0	20.1	12.4	68.5	5.6	12.4
TtRnaseH	ΔH_G	120.0	1.8	86.0	25.2	10.5	68.5	5.1	10.5
TtRnaseH	ΔC_p	131.0	2.7	86.0	40.6	8.5	68.5	5.2	7.4
TtRnaseH	$\Delta H_G, \Delta C_p$	120.0	2.7	86.0	44.2	7.1	68.5	4.7	5.5

^a EcRnaseH stands for *E. coli* RnaseH, and TtRnaseH stands for *T. thermophilus* Rnase H. The thermodynamic data are for the cysteine-free variants of EcRnaseH and TtRnaseH (35, 36).

particular DSC scan at a pH of 7.0 (Table 1 in ref 32). Nevertheless, because our database is quite small, we cannot rule out other ways of achieving a higher $\Delta G(T_S)$ in thermophilic proteins. Furthermore, we cannot rule out other routes to achieve a higher T_G than by increasing $\Delta G(T_S)$. However, it appears likely that this route is preferred.

To understand the relative contributions of increasing ΔH_G and lowering ΔC_p toward an increase in $\Delta G(T_S)$, we have simulated hypothetical protein stability curves for one TM family, the cysteine-free variants of Rnase H from *E. coli* (EcRnaseH) and from *T. thermophilus* (TtRnaseH) (Figure

4). These hypothetical protein stability curves were obtained by interchanging the ΔC_p and ΔH_G values between TtRnaseH and EcRnaseH. Table 4 lists the thermodynamic parameters obtained from these hypothetical stability curves. In Figure 4, the protein stability curves derived from experimental data are shown in solid lines. The protein stability curve for TtRnaseH is shown in red and that for EcRnaseH in blue. Interchanging only the ΔH_G values between EcRnaseH and TtRnaseH increases the stability of EcRnaseH and decreases that of TtRnaseH (dotted curves). Similarly, interchanging only the ΔC_p values between EcRnaseH and TtRnaseH also

increases the stability of EcRnaseH and decreases the stability of TtRnaseH (curves traced by “+” marks). The effect in this case is larger (Table 4). However, simultaneously interchanging ΔH_G and ΔC_p between EcRnaseH and TtRnaseH has the greatest effect (dashed curves). Because of this simultaneous interchange, the stability of EcRnaseH at room temperature increases to become comparable to that of TtRnaseH (Table 4). For EcRnaseH, $\Delta G(T_L)$ at 37 °C is 6.8 kcal/mol. Exchanging ΔH_G elevates $\Delta G(T_L)$ from 6.8 to 7.8 kcal/mol. Exchanging ΔC_p elevates $\Delta G(T_L)$ from 6.8 to 8.0 kcal/mol. Exchanging both yields 8.9 kcal/mol (Table 4). However, this increase in protein stability may impair function, unless the organism adapts to a higher temperature.

DISCUSSION

The heat capacity change, ΔC_p , between the native and the denatured states of a protein remains constant in the range 20–80 °C. At those temperatures outside this range, a decrease in ΔC_p is temperature dependent (2). Hence, the assumption of a constant ΔC_p , used in plotting protein stability curves (Figures 1 and 2), is an approximation. The effect of this approximation may be larger for thermophiles than for mesophiles. However, within this approximation, our results clearly show that the higher temperature resistance of the thermophilic proteins is due to their greater thermodynamic stability.

At different temperatures, different factors contribute dominantly to protein stability. To understand the effect of temperature on the native state of the proteins, it is convenient to use a two-state model of the water structure for the surrounding water molecules. Using the water structure with “regular” hexagonal enthalpically favored ice prevailing at low temperatures and increasingly fluctuating the high-density entropically favored liquid water structural forms at higher temperatures, we are able to interpret and predict these results microscopically. The hydrophobic effect does not exist in the cold, consistent with the predicted lack of correlation between ΔC_p and T'_G . However, ΔH_G and ΔS_G are predicted to correlate with T_G . These predictions have been corroborated by the observed correlations in our database.

Ideally, the correlation analysis should be applied to the data on protein stability determined under identical conditions. However, in practice, different laboratories use different experimental procedures and reagents in protein thermal studies. Here, our aim is to collect the thermodynamic data on different proteins determined in as similar conditions as possible. The thermodynamic data on most proteins used in this study are around a neutral pH. However, for RnaseH and archaeal histones, the data used here are at pH values of 5.5 and 5.0, respectively (detailed in the Materials and Methods section). We retain these proteins because their removal would have substantially depleted the already sparse thermodynamic data currently available on homologous thermophilic and mesophilic proteins. The effect of the variation in experimental conditions on the observed values of the square of the linear correlation coefficient (r^2) for various parameter pairs is unknown. However, we do not expect this effect to be large. We use r^2 instead of r to identify correlations between various thermodynamic parameter pairs. This further minimizes the chances of obtain-

ing spurious results. The t test provides further indications if the correlations between the different thermodynamic parameters observed here are likely to be observed in proteins in general.

There are several possible reasons behind the observed lack of correlation between ΔC_p and T_G . First, the correlation may be seen in a larger database of proteins but is not observed in our study because of sparse data. Consequently, we have performed a t test, resulting in the *acceptance* of the null hypothesis (see the Materials and Methods section and Table 2). The two parameters appear uncorrelated in proteins. Recently, we have compiled a database of two-state folding proteins which show reversible two-state folding \rightleftharpoons unfolding transitions at or near a physiological pH. A preliminary examination of this larger database also does not indicate a correlation between ΔC_p and T_G . Second, the correlation is not seen because of the inaccuracies in the experimental determination of ΔC_p . Because ΔC_p is directly related to ΔASA (9–11), we have estimated ΔASA from the empirical relationship between ΔASA and N_{res} , the number of residues, given in ref 9. We find that ΔASA also does not correlate with T_G . Similarly, there is no correlation between the melting temperature (T_G) and protein size (N_{res}). Hence, it appears that the lack of correlation between ΔC_p and T_G is biologically meaningful.

Can we interpret the observed correlations (or lack of correlations) among various thermodynamic parameters in terms of sequence and structural differences among homologous thermophilic and mesophilic proteins? Recently, we have analyzed the contribution of various factors (hydrophobicity, compactness, proline content, disulfide bridges, residue composition, secondary structure content, surface areas, insertions and deletions, oligomerization, and electrostatic interactions such as salt bridges and hydrogen bonds) in 18 nonredundant families containing homologous thermophilic and mesophilic proteins (23). The parameters that determine the overall protein fold have similar values for thermophilic and mesophilic proteins. Hence, the thermophilic and mesophilic proteins have percentages similar to those of a nonpolar surface buried in the core (hydrophobicities) and compactness (atomic packing), indicating a similar extent of hydrophobic contribution toward the stability of these proteins. The distribution of buried and exposed polar and nonpolar surface areas is quite uniform in proteins and does not change among thermophiles and mesophiles (23). The occurrence of main chain–main chain hydrogen bonds that define protein secondary structure is also similar between thermophiles and mesophiles. Even though these parameters are important contributors toward the overall protein folding and stability, their contribution toward protein stability *differentials* between homologous thermophiles and mesophiles appears to be small. These observations are consistent with the noncorrelation of T_G with ΔC_p , ΔASA , or N_{res} .

Despite the high sequence identity among thermophiles and mesophiles, their amino acid distributions are significantly different. The thermophilic proteins appear to favor amino acid residues with larger side chains and to avoid thermolabile residues. An increase in electrostatic interactions, salt bridges, and side chain–side chain hydrogen bonds in thermophiles as compared to their homologous mesophiles is the most consistent trend across the 18 nonredundant families of homologous thermophilic and mesophilic proteins

(23). These observations are consistent with the correlations among ΔH_G , ΔS_G , $\Delta G(T_S)$, and T_G reported here.

Analyses based on protein sequences from complete genomes of thermophilic and hyperthermophilic organisms and comparisons of the sequence and structural properties of homologous thermophilic and mesophilic proteins have consistently indicated a significant increase in the proportion of charged residues for the thermophiles (50, 51). The improvement in electrostatics for thermophiles may be reflected as an alleviation of electrostatic repulsions, an increased occurrence of ion pairs and their networks, and a geometrical optimization of charged residue positions to yield a favorable energetic contribution toward protein stability. Different protein families may optimize these factors differently. In the following paragraphs, we present evidence supporting the role of electrostatic interactions in enhancing the stability of thermophilic proteins in the T/M families in our database.

We have performed a comprehensive sequence and structure comparison of Ribonuclease H from *T. thermophilus* (TtRnaseH) and from *E. coli* (EcRnaseH). TtRnaseH and EcRnaseH have similar atomic packing and extents of nonspecific interactions. However, TtRnaseH has a greater proportion of charged residues and more close-range electrostatic interactions, such as ion pairs and salt bridges and their networks. The sequence alignment of EcRnaseH and TtRnaseH indicates eight positions where the apolar residues in EcRnaseH have been replaced by charged residues in TtRnaseH. We have computed the electrostatic free-energy contributions ($\Delta\Delta G_{\text{elec}}$) for six of these charged residues to the stability of TtRnaseH, using a method described previously (52). The remaining two substitutions are at the N and C termini of TtRnaseH, and the atomic coordinates for the adjoining terminal residues are missing in the crystal structure of TtRnaseH. Five of the six apolar-to-charged substitutions have a stabilizing electrostatic free-energy contribution (data not shown).

Using site-directed mutagenesis, Li et al. (53) have attempted to interpret the differences in the thermodynamic stabilities of the hyperthermophilic histone rHMfB and its mesophilic homologue rHFoB in terms of their sequence and structural differences. Their results indicate that improved hydrophobic interactions in the histone-dimer core, alleviation of electrostatic repulsion, and formation of additional ion pairs at the dimer interface are responsible for the higher thermostability of rHMfB.

Frankenberg et al. (54) have built a model of *T. maritima* cold-shock protein (CspTm) by homology modeling. The model indicates an increased number of surface salt bridges in CspTm. Mueller et al. (55) have solved the X-ray structure of a cold-shock protein from the thermophile *Bacillus caldolyticus* (CspBc) and compared it with that of CspBs, *B. subtilis* cold-shock protein. The distribution of surface charges was found to be different and overall favorable in CspBc. The amino acid sequences of CspBc and CspBs differ at 12 positions. Perl et al. (56) have produced 12 variants of CspBc, each containing an amino acid substitution in CspBc of the residue in the corresponding position in CspBs. They have reported that only two surface-exposed residues, Arg3 and Leu66, are responsible for the increase in the stability of the thermophilic protein.

de Bakker et al. (57) have performed molecular dynamic simulations of Sac7d from the hyperthermophile *Sulfolobus acidocaldarius* at 300, 360, and 550 K. They concluded that salt bridges contribute favorably toward protein stability at elevated temperatures.

Recently, we have studied the energetic contribution of electrostatic interactions toward protein stability (24, 58). Salt bridges and ion pairs may be stabilizing or destabilizing to the protein structure depending upon a number of factors (58). Hydration free energies of amino acids decrease at high temperatures because of a decrease in the dielectric constant of solvent water (59). This indicates a reduced energy penalty for desolvation of charged residues in folded proteins at high temperatures. The optimization of electrostatic interactions in hyperthermophilic proteins results in their greater electrostatic contribution (25). Close-range electrostatic interactions, such as ion pairs and salt bridges and their networks, contribute favorably to the stability of *Pyrococcus furiosus* glutamate dehydrogenase (24). The alleviation of electrostatic repulsion and optimization of long-range electrostatic interactions by mutating charged residues on the protein surface has been shown to improve protein stability (52, 60, 61). Increased and improved electrostatic interactions should not be taken to imply greater rigidity for thermophilic proteins. Proteins are flexible at the optimum temperatures of their source organisms (62).

CONCLUSIONS

Proteins show good correlations among the melting temperature, maximal protein stability, and enthalpy and entropy changes at the melting temperature. These correlations indicate that higher melting temperatures in thermophilic proteins are due to an upshift and broadening of their stability curves. Specific (e.g., electrostatic) interactions found in thermophiles but absent in mesophiles are important contributors to the higher stability of the thermophiles. Homologous thermophilic and mesophilic proteins have similar stabilities at their source-organism living temperatures. To be functional, the protein should not be too stable; hence, their working stabilities (at the source-organism living temperatures) are lower than their maximal stabilities. Because the stability curves are roughly symmetric (skewed parabolas), for most proteins, a higher melting temperature also indicates a lower cold-denaturation temperature. This implies that, to obtain lower cold-denaturation temperatures, the key is, again, specific interactions. These observations have direct implications to protein design, indicating that, to achieve higher stability, specific and largely electrostatic interactions need to be engineered into the proteins.

ACKNOWLEDGMENT

We thank Drs. Buyong Ba, Yuk Yin Sham, Neeti Sinha, Kannan Gunasekaran, David Zanuy, and, in particular, Jacob V. Maizel for numerous helpful discussions.

APPENDIX

For a protein or a protein domain which (i) folds in a reversible two-state ($N \rightleftharpoons D$) manner, (ii) is stable over a temperature range, and (iii) has a constant (greater than zero) heat capacity change in this range, the Gibbs–Helmholtz equation can be used to plot its stability curve (1, 2).

$$\Delta G(T) = \Delta H_G(1 - T/T_G) - \Delta C_p[(T_G - T) + T \ln(T/T_G)] \quad (1)$$

where $\Delta G(T)$ is the Gibbs free-energy change between the denatured (D) and the native (N) states of the protein at a given temperature T . Three thermodynamic parameters are needed to plot a protein stability curve: the melting temperature (T_G), the change in enthalpy at T_G (ΔH_G), and the change in heat capacity (ΔC_p). These parameters can be determined by experimental means. The following quantities are computed from the protein stability curves (1). The slope of the protein stability curve is given by

$$\partial \Delta G(T) / \partial T = -\Delta H_G / T_G - \Delta C_p \ln(T/T_G) \quad (2)$$

Hence, at $T = T_G$,

$$\partial \Delta G(T) / \partial T = -\Delta H_G / T_G = -\Delta S_G \quad (3)$$

where ΔS_G is the entropy change between native and denatured states of the protein at the melting temperature. The curvature of the protein stability curve at temperature T is given by

$$\partial^2 \Delta G(T) / \partial T^2 = -\Delta C_p / T \quad (4)$$

Hence, at $T = T_G$,

$$\partial^2 \Delta G(T) / \partial T^2 = -\Delta C_p / T_G \quad (5)$$

The temperature of maximal protein stability, T_S , is also the temperature at which the entropy change between the native and denatured states of the protein, ΔS , is zero. It is given by

$$T_S = T_G \exp[-\Delta H_G / (T_G \Delta C_p)] \quad (6)$$

The maximal protein stability change, $\Delta G(T_S)$ (i.e., the free-energy change at T_S) is given by

$$\Delta G(T_S) = \Delta H(T_S) = \Delta H_G - (T_G - T_S) \Delta C_p \quad (7)$$

These parameters and equations provide the macroscopic thermodynamic description of proteins.

REFERENCES

- Becktel, W., and Schellman, J. A. (1987) *Biopolymers* 26, 1859–1877.
- Privalov, P. L. (1990) *Crit. Rev. Biochem. Mol. Biol.* 25, 281–305.
- Pace, C. N., and Shaw, K. L. (2000) *Proteins* 41, S4, 1–7.
- Shortle, D. (1995) *Adv. Protein Chem.* 46, 217–247.
- Wrabl, J., and Shortle, D. (1999) *Nat. Struct. Biol.* 6, 876–882.
- Griko, Y. V. (2000) *J. Mol. Biol.* 297, 1259–1268.
- Hallerbach, B., and Hinz, H. J. (1999) *Biophys. Chem.* 76, 219–227.
- Brandts, J. F. (1964) *J. Am. Chem. Soc.* 86, 4291–4314.
- Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) *Protein Sci.* 4, 2138–2148.
- Edgcomb, S. P., and Murphy, K. P. (2000) *Curr. Opin. Biotechnol.* 11, 62–66.
- Makhatadze, G. I., and Privalov, P. L. (1990) *J. Mol. Biol.* 213, 375–384.
- Kauzmann, W. (1959) *Adv. Protein Chem.* 14, 1–63.
- Dill, K. A. (1990) *Biochemistry* 31, 7134–7155.
- Schellman, J. A. (1997) *Biophys. J.* 73, 2960–2964.
- Privalov, P. L., and Gill, S. J. (1988) *Adv. Protein Chem.* 39, 191–234.
- Vedamuthu, M., Singh, S., and Robinson, G. W. (1994) *J. Phys. Chem.* 98, 2222–2230.
- Robinson, G. W., and Cho, C. H. (1999) *Biophys. J.* 77, 3311–3318.
- Ruelle, P., and Kesselring, U. W. (1998) *J. Pharm. Sci.* 87, 987–997.
- Eliezer, D., and Wright, P. E. (1996) *J. Mol. Biol.* 263, 531–538.
- Privalov, P. L. (1996) *J. Mol. Biol.* 6, 860–872.
- Fink, A. L., Oberg, K. A., and Sehadri S. (1997) *Folding Des.* 3, 19–25.
- Polverino de Laureto, P., Scaramella, E., Frigo, M., Wondrich, F. G., De Filippis, V., Zamboni, M., and Fontana, A. (1999) *Protein Sci.* 8, 2290–2303.
- Kumar, S., Tsai, C. J., and Nussinov, R. (2000) *Protein Eng.* 13, 179–191.
- Kumar, S., Ma, B., Tsai, C. J., and Nussinov, R. (2000) *Proteins* 38, 368–383.
- Xiao, L., and Honig, B. (1999) *J. Mol. Biol.* 289, 1435–1444.
- Gromiha, M. M., An, J., Kono, H., Oobatake, M., Uedaira, H., Prabakaran, P., and Sarai, A. (2000) *Nucleic Acids Res.* 28, 283–285.
- Li, W., Grayling, R. A., Sandman, K., Edmondson, S., Shriver, J. W., and Reeve, J. N. (1998) *Biochemistry* 37, 10563–10572.
- Knapp, S., Karshikoff, A., Berndt, K. D., Christova, P., Atanasov, B., and Ladenstein R. (1996) *J. Mol. Biol.* 264, 1132–1144.
- Knapp, S., Mattson, P. T., Christova, P., Berndt, K. D., Karshikoff, A., Vihinen, M., Smith, C. I. E., and Ladenstein, R. (1998) *Proteins* 31, 309–319.
- Filimonov, V. V., Azuaga, A. I., Viguera, A. R., Serrano, L., and Mateo, P. L. (1999) *Biophys. Chem.* 77, 195–208.
- McCrary, B. S., Edmondson, S. P., and Shriver, J. W. (1996) *J. Mol. Biol.* 264, 784–805.
- Wassenberg, D., Welker, C., and Jaenicke, R. (1999) *J. Mol. Biol.* 289, 187–193.
- Petrosian, S. A., and Makhatadze, G. I. (2000) *Protein Sci.* 9, 387–394.
- Pfeil, W. (1998) *Protein stability and folding: A collection of thermodynamic data*, Springer-Verlag, Heidelberg, Germany.
- Hollien, J., and Marqusee, S. (1999) *Proc. Natl. Acad. Sci. U.S.A.* 96, 13674–13678.
- Hollien, J., and Marqusee, S. (1999) *Biochemistry* 38, 3831–3836.
- Beadle, B. M., Baase, W. A., Wilson, D. B., Gilkes, N. R., and Shoichet, B. K. (1999) *Biochemistry* 38, 2570–2576.
- Chiti, F., van Nuland, N. A. J., Taddei, N., Magherini, F., Stefani, M., Ramponi, G., and Dobson, C. M. (1998) *Biochemistry* 37, 1447–1455.
- Nicholson, E. M., and Scholtz, J. M. (1996) *Biochemistry* 35, 11369–11378.
- Taddei, N., Chiti, F., Paoli, P., Fiaschi, T., Bucciantini, M., Stefani, M., Dobson, C. M., and Ramponi, G. (1999) *Biochemistry* 38, 2135–2142.
- Pace, C. N., Hebert, E. J., Shaw, K. L., Schell, D., Both, V., Krajcikova, D., Sevcik, J., Wilson, K. S., Dauter, Z., Hartley, R. W., and Grimsley, G. R. (1998) *J. Mol. Biol.* 279, 271–286.
- Privalov, P. L., and Khechinashvili, N. (1974) *J. Mol. Biol.* 86, 665–684.
- Privalov, P. L. (1979) *Adv. Protein Chem.* 33, 167–241.
- Pace, C. N., Grimsley, G. R., Thomas, S., and Makhatadze, G. I. (1999) *Protein Sci.* 8, 1500–1504.
- Spiegel, M. R. (1980) *Theory and problems of probability and statistics* SI (metric) ed., Schaum's Outline Series, McGraw-Hill, New York.
- Alexander, P., Fahnestock, S., Lee, T., Orban, J., and Bryan, P. (1992) *Biochemistry* 31, 3597–3603.
- Rees, D. C., and Adams, M. W. W. (1995) *Structure* 3, 251–254.

48. Rees, D. C., and Robertson, A. D. (2001) *Protein Sci.* 10, 1187–1194.
49. Kumar, S., Ma, B., Tsai, C. J., Sinha, N., and Nussinov, R. (2000) *Protein Sci.* 9, 10–19.
50. Kumar, S., and Nussinov, R. (2001) *Cell. Mol. Life Sci.*, in press.
51. Cambillau, C., and Claverie, J.-M. (2000) *J. Biol. Chem.* 275, 32383–32386.
52. Spector, S., Wang, M., Carp, S. A., Robblee, J., Hendsch, Z. S., Fairman, R., Tidor, B., and Raleigh, D. P. (2000) *Biochemistry* 39, 872–879.
53. Li, W.-T., Shriver, J. W., and Reeve, J. N. (2000) *J. Bacteriol.* 182, 812–817.
54. Frankenberg, N., Welker, C., and Jaenicke, R. (1999) *FEBS Lett.* 454, 299–302.
55. Mueller, U., Perl, D., Schmid, F. X., and Heinemann, U. (2000) *J. Mol. Biol.* 297, 975–988.
56. Perl, D., Mueller, U., Heinemann, U., and Schmid, F. X. (2000) *Nat. Struct. Biol.* 7, 380–383.
57. de Bakker, P. I. W., Hunenberger, P. H., and McCammon, J. A. (1999) *J. Mol. Biol.* 285, 1811–1830.
58. Kumar, S., and Nussinov, R. (1999) *J. Mol. Biol.* 293, 1241–1255.
59. Elcock, A. H. (1998) *J. Mol. Biol.* 284, 489–502.
60. Grimsley, G. R., Shaw, K. L., Fee, L. R., and Alston, R. W. Huyghues-Despointes, B. M., Thurlkil, R. L., Scholtz, J. M., and Pace, C. N. (1999) *Protein Sci.* 8, 1843–1849.
61. Loladze, V. V., Ibarra-olero, B., Sanchez-Ruiz, J. M., and Makhataдзе, G. I. (1999) *Biochemistry* 38, 16419–16423.
62. Jaenicke, R. (2000) *Proc. Natl. Acad. Sci. U.S.A.* 97, 2962–2964.

BI0106383